

## University of Groningen

### Practice-inspired contributions to inventory theory

Prak, Derk Rutger Jordi

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2019

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Prak, D. R. J. (2019). *Practice-inspired contributions to inventory theory*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen, SOM research school.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## CHAPTER 2

# On the calculation of safety stocks when demand is forecasted

---

**Abstract.** *The inventory control literature generally assumes that the demand distribution and all its parameters are known. In practical applications it is often suggested to estimate the demand variance either directly or based on the one-period ahead forecast errors. The variance of the lead time demand, essential for safety stock calculations, is then obtained by multiplying the estimated per-period demand variance by the length of the lead time. However, this is flawed, since forecast errors for different periods of the lead time are positively correlated, even if the demand process itself does not show (process) auto-correlation. As a result these traditional procedures lead to safety stocks that are too low. This paper presents corrected lead time demand variance expressions and reorder levels for inventory systems with a constant lead time where demand fluctuates around a constant level. Firstly, we derive the exact lead time forecast error of mean demand conditional on the true demand variance. Secondly, we derive for normally distributed demand the correct reorder level under uncertainty of both the demand mean and variance. We show how the results can be implemented in inventory models, and particularly discuss batch ordering policies combined with moving average and exponential smoothing forecasts. We find that traditional approaches can lead to safety stocks that are up to 30% too low and service levels that are up to 10% below the target.*

---

This chapter is based on Prak et al. (2017b): Prak, D.R.J., R.H. Teunter, A.A. Syntetos. 2017b. On the calculation of safety stocks when demand is forecasted. *European Journal of Operational Research* **256**(2) 454–461.

## 2.1 Introduction

Inventory control models in the research literature and in textbooks typically assume that the demand distribution and all its parameters are known. However, in practice such information is not available, and future demands have to be forecasted based on historical observations. No forecasting procedure produces perfect forecasts. For unbiased forecasting procedures, the forecast errors of (mean) demand can be decomposed into the fluctuation of the demand around its true mean, and the fluctuation of the estimates around this true mean. The latter term is positively correlated over all future periods, and for i.i.d. demand the latter term is exactly the same for all future periods. Therefore, future forecast errors are correlated, even if the demand process itself shows no auto-correlation. This issue of correlation and more generally the integration of forecasting and inventory decision making is understudied, and results in safety stocks that are set too low, and service targets that are not achieved.

We discuss the stationary (i.i.d.) demand model in combination with a constant lead time, which is used as a starting point in most inventory control textbooks. Since demands are independent over time, the lead time demand variance for this demand model is easily obtained by multiplying the lead time with the per-period demand variance. For normally distributed demand, it is well known that the safety stock needed to obtain a certain service level is obtained by multiplying the lead time demand standard deviation with some safety factor. In practice, however, these results can never be directly applied as the mean and variance of the per-period demand are not known but need to be forecasted. Surprisingly, some inventory textbooks do not mention forecasting at all (Zipkin, 2000; Muckstadt and Sapra, 2010) or do not discuss the incorporation of forecasts and their errors in the decision models (Waters, 2012). Some inventory textbooks, such as Silver et al. (2017) and Axsäter (2015), do indicate that parameters have to be estimated in practical applications, and that the resulting forecast error should be taken into account. To the best of our knowledge, all the books in this latter category suggest to use the variance of the per-period demand forecast errors (e.g. estimated via Mean Squared Error, MSE, or Mean Absolute Deviation, MAD). Although this correctly captures the per-period forecast error in the demand mean, it ignores correlation of forecast errors for the different periods

of a lead time (interval), and it furthermore ignores the forecast error in the demand variance.

In this paper, we consider the practical situation where the mean  $\mu$  and variance  $\sigma^2$  of demand are unknown, and there is a positive lead time. We provide corrected expressions for the variance of the lead time forecast error for any forecasting technique under the stationary demand model, and highlight the corrections needed for two of the most widely used forecasting methods: Simple Moving Average (SMA) and Single Exponential Smoothing (SES). We furthermore derive corrected expressions for the reorder level under a cycle service level constraint. Although the procedure is similar for other service level definitions or cost equations, the cycle service definition allows us to focus solely on the reorder level and thus on safety stocks<sup>1</sup>. The analysis consists of two stages. First, we show how to correctly deal with forecast errors in the demand mean, conditional on the true demand variance. Second, we discuss the simultaneous treatment of forecast errors in both the mean and variance of the demand. In the second stage, we focus on normally distributed demand, which is often used in research and practice, see e.g. Zeng and Hayya (1999), Strijbosch and Moors (2006), and Axsäter (2013). We show that the corrected procedures typically lead to considerably higher safety stocks, and it will become obvious from our analysis that the same type of correction is needed for other demand processes, forecasting procedures, and inventory models. In fact, for any inventory control application where demand is forecasted based on a finite number of demand observations, the traditional approaches for determining safety stocks suffer from the same problem.

The remainder of this paper is structured as follows. In Section 2.2, we discuss the related literature. Then, in Section 2.3, we formally introduce the stationary demand model and show how to correctly include forecast errors in the estimation of the mean of per-period demand for any forecasting procedure. In Section 2.4 we discuss the correction factors for safety stocks. Section 2.5 discusses how to correct for uncertainty of both the mean and the variance of the demand. In Section 2.6 we use the obtained results to derive reorder levels for batch order policies. Section 2.7

---

<sup>1</sup>We assume throughout this paper (in line with the majority of the literature) that an order can be placed when the inventory equals exactly the reorder level, and that there is no undershoot due to demand arriving in batches or periodic review. If such an undershoot can occur, then the achieved service level will be below its target.

provides a numerical study with an evaluation of reorder and service levels corresponding to traditional approaches and the corrected approach. Finally, Section 2.8 concludes.

## 2.2 Related literature

The fact that lead time forecast errors may be correlated over time has been pointed out by other researchers. Silver et al. (2017) already noted that the relationship between the forecast error during the lead time and that during the forecast interval *“depends in a complicated fashion on the specific underlying demand model, the forecast updating procedure and the values of the smoothing constant used.”* Subsequently they argued that the relationship concerned should be established empirically. Beutel and Minner (2012) discuss an integrated framework of least squares demand forecasting and inventory decision making. They find that if demand parameters are replaced by estimates, and the estimates are based on few observations, then actual service levels will significantly undershoot their targets.

Other authors discuss the related issue that the actual demand process may be more variable than assumed. Johnston and Harrison (1986) and Graves (1999) show that fluctuations in the mean demand level lead to additional uncertainty compared to the stationary demand model, and they propose how the variance of lead time demand should be adjusted accordingly. Similarly, fluctuations in the variance of per-period demand are treated by Zhang (2007). These researchers rightly suggest that if there is more variance that should be taken into account, then ignoring it will lead to an underestimation of the safety stocks needed to sustain a certain service level performance. However, these arguments and results are fundamentally different from ours. We show that forecast errors are correlated over the lead time even if demands are not. Therefore, the traditional safety stock calculations that ignore this are flawed even if the demand model is specified correctly and the best (minimum variance unbiased) estimator is used.

A general approach to inventory decision making under unknown demand parameters is the Bayesian approach, which was pioneered by Dvoretzky et al. (1952b), Scarf (1959), Karlin (1960), and Iglehart (1964). Applications of Bayesian estimation in dynamic programming set-ups are given by e.g. Azoury (1985) and Lariviere and

Porteus (1999). It has been found that exact, optimal order quantities are numerically hard to derive. This led to the development of heuristics, especially myopic policies that essentially reduce multi-period problems to single-period problems, and typically assume a negligible lead time. See e.g. Chen (2010). Numerical Bayesian studies mainly involve single-parameter demand distributions, or, if the distribution has more than one parameter, impose that one or more parameters are known. Azoury and Miyaoka (2009) and Chen (2010) use a normal distribution of which the variance is assumed to be known, and Rajashree Kamath and Pakkala (2002) use a log-normal distribution with a known variance. This restriction is theoretically convenient, because it ensures that the posterior demand distribution is again (log)normal, but it is practically incorrect, as the variance has to be estimated as well.

A different approach is given by Hayes (1969), who introduces the method of minimizing the Expected Total Operating Cost, which follows from the sampling distribution of the parameter estimators. Related work for the newsvendor model is that by Kevork (2010) and Akcay et al. (2011). Ritchken and Sankar (1984) derive the optimal reorder level for a single period inventory system under normally distributed demand with unknown mean and variance. They derive a consistent estimator of quantiles of the normal distribution when the mean and variance are unknown. Then, they choose as the reorder level the quantile corresponding to the required service level. Janssen et al. (2009) seek the optimal 'upward bias' in the traditional stock level calculation via simulation. Similarly, Strijbosch and Moors (2005) consider a batch ordering  $(r, Q)$  policy under normally distributed demand with unknown parameters and zero lead time. They take the reorder level  $r$  as fixed and optimize the batch size  $Q$  via simulation. Like the literature on Bayesian inventory modeling, the focus of the mentioned non-Bayesian approaches is also on single-period models and/or models with zero lead time.

Some authors do take into account the possibility of a positive lead time, but focus on stochastic lead times and assume that the demand variance is known. An example is Eppen and Martin (1988), who discuss how to derive the forecast error in the lead time demand when the demand mean is estimated by single exponential smoothing. They do not discuss the relationship with actual inventory decision making. Namit and Chen (2007) follow a similar approach, but consider a more general framework of possibly auto-correlated demand and ARMA optimal forecasting.

Although they do discuss the relationship between forecasting and determining the reorder level, they also ignore the uncertainty in the demand variance.

Silver and Rahnema (1986, 1987) are, to the best of our knowledge, the only authors who discuss the combination of normally distributed demand with unknown mean and variance and a positive fixed lead time. They derive a factor to adjust the traditional reorder level upward. This factor has to be computed by solving a non-linear equation involving an integral expression. They find that the required adjustment factor is substantial under many parameter settings. Although their results are derived for a model with holding and shortage costs, these results can be translated to equivalent service level models. In Section 2.7 we benchmark their complex, approximate method to our simpler, exact approach.

The practical situation where the demand mean and variance are unknown and a positive lead time is present, remains ill-studied in the literature. Since no closed-form optimal reorder level exists yet in this case, the main approach used in practical applications remains to multiply either the directly estimated standard deviation of demand, or the square root of the MSE, by the length of the lead time interval. This paper does provide a simple, closed-form expression for the corrected reorder levels under uncertainty of both the mean and the variance of the demand, and furthermore discusses the magnitude of the flaw of the traditional approaches.

## 2.3 Corrected variance of the lead time demand when $\mu$ is unknown

We consider a stationary demand process; that is, demands  $Y_t$  ( $t = 1, 2, \dots$ ) are independent over time and fluctuate around a constant mean  $\mu$  with constant variance  $\sigma^2$ :  $Y_t \stackrel{\text{i.i.d.}}{\sim} (\mu, \sigma^2)$ . At this point we assume that  $\mu$  is unknown, whereas  $\sigma^2$  is known. We relax the latter assumption in Section 2.5. Since demands are independent over time  $t$  and have the same mean, we assume that the forecasting procedure at the end of time  $N$  produces the same forecast  $\hat{Y}_N$  for all future periods  $N+k$ ,  $k = 1, 2, \dots$ . Obviously,  $\hat{Y}_N$  is an estimator of  $\mu$ . This forecast  $\hat{Y}_N$  is a function of demands  $Y_1, \dots, Y_N$ , although not necessarily all these historical demands are utilized in the forecast. The naive forecast, for instance, simply uses the last observation  $Y_N$ , whereas the mini-

imum variance unbiased estimator averages all past demand observations  $Y_1, \dots, Y_N$ . Furthermore, there is a constant lead time  $L$ .

For any  $N$  and  $k$  ( $N, k = 1, 2, \dots$ ), denote the forecast error of demand at time  $N + k$  based on the forecast made at the end of time  $N$ , by  $\epsilon_{N+k}$ . That is, define

$$\epsilon_{N+k} = Y_{N+k} - \hat{Y}_N = (Y_{N+k} - \mu) + (\mu - \hat{Y}_N).$$

Both  $Y_{N+k}$  and  $\hat{Y}_N$  are random variables, the difference being that at the end of time  $N$ ,  $Y_{N+k}$  is still unobserved, although the forecast  $\hat{Y}_N$  (of mean  $\mu$ ) is known. The lead time forecast error can be decomposed as follows:

$$\begin{aligned} \sum_{k=1}^L \epsilon_{N+k} &= \sum_{k=1}^L Y_{N+k} - L\hat{Y}_N \\ &= \sum_{k=1}^L (Y_{N+k} - \mu) + \sum_{k=1}^L (\mu - \hat{Y}_N) \\ &= \sum_{k=1}^L (Y_{N+k} - \mu) + L(\mu - \hat{Y}_N). \end{aligned}$$

The former term is the deviation of lead time demand from its true mean and the latter term represents the (correlated) deviations of the estimate  $\hat{Y}_N$  around the mean  $\mu$  of the per-period demand.

For ease of exposition, and in line with the inventory control literature, we will restrict our attention to unbiased estimators. That is, the expectation of  $\hat{Y}_N$  equals  $\mu$ , which clearly implies that the expectation of the second term of the lead time forecast error is zero. Obviously, the same holds for the first term. Moreover, as the second term is based on past demands, and the first on future demands, and demands are assumed to be independent over time, the two terms are independent. Therefore, the variance of the lead time forecast error is given by

$$\text{Var} \left( \sum_{k=1}^L \epsilon_{N+k} \right) = L\sigma^2 + L^2 \text{Var}(\hat{Y}_N), \quad (2.1)$$

where  $\text{Var}(\hat{Y}_N)$  denotes the variance of  $\hat{Y}_N$ . We remark that (2.1) is in accordance with



Strijbosch et al. (2000) and Syntetos et al. (2005), who offer different and arguably less insightful proofs. Furthermore, (2.1) shows that *only* if one can perfectly predict  $\mu$ , then the traditional lead time demand variance corresponds to the lead time forecast error. A key observation is that  $\text{Var}(\hat{Y}_N)$  is multiplied by  $L^2$  in the variance of the lead time forecast error. This is a consequence of the fact that the forecast error  $(\mu - \hat{Y}_N)$  is the same for each period of the lead time interval, implying positive covariance. It is this covariance that is ignored by standard inventory theory and software.

Some textbooks do suggest to replace the per-period demand variance by the per-period forecast error variance (estimated via MSE or MAD), which we will refer to as the ‘per-period MSE approach’. This approach calculates the lead time forecast error as the sum of the per-period forecast errors, which implies that their variances are added and their correlations are ignored. This leads to

$$\sum_{k=1}^L \text{Var}(\epsilon_{N+k}) = \sum_{k=1}^L \text{Var}(Y_{N+k} - \hat{Y}_N) = L \left( \sigma^2 + \text{Var}(\hat{Y}_N) \right) = L\sigma^2 + L\text{Var}(\hat{Y}_N). \quad (2.2)$$

The second alternative approach that is still frequently applied is ‘direct substitution’ of the demand variance  $\sigma^2$  into the decision model. In this view, the variance of the demand during the lead time is the sum of the  $L$  variances of demand during a single period, leading to  $L\sigma^2$  as the variance of the lead time demand (forecast error).

The calculations above can be translated into a corrected specification of the lead time demand distribution. That distribution can be used in an inventory model as a substitute for the true lead time demand distribution, in case demand is forecasted. The corrected approach specifies the lead time demand distribution as a normal distribution with mean  $L\hat{Y}_n$  and variance  $L\sigma^2 + L^2\text{Var}(\hat{Y}_N)$ . For completeness, the MSE approach specifies a normal distribution with mean  $L\hat{Y}_n$  and variance  $L \left( \sigma^2 + \text{Var}(\hat{Y}_N) \right)$ , and the traditional demand variance approach specifies a normal distribution with mean  $L\hat{Y}_n$  and variance  $L\sigma^2$ . This derivation is key to generalizing the results from this paper to different inventory models.

## 2.4 Correction factors for safety stocks

In this section we discuss the implications of ignoring error auto-correlation for different forecasting methods. We first consider the SMA forecasting method. Observe that a special case of this method is taking the average over all past demand observations, which is the minimum variance unbiased estimator for  $\mu$  in the stationary demand model. We will show that also in that special case, safety stock calculations need to be corrected. Then, we develop the correction factor for SES, another forecasting technique that is widely applied in practice.

We provide exact expressions under the assumption that the demand variance  $\sigma^2$  is known. In Section 2.5 we relax this assumption. We realize that neither SMA (except for the above mentioned special case) nor SES are the minimum variance unbiased estimators for a level demand process. See also Newbold and Bos (1989). However, in practice it is typically observed that the mean of the demand process only remains constant for a limited time. It is unknown whether the mean demand is constant and if so, since when. This explains the popularity of SMA and SES in practice and their good performance in forecasting competitions (Gardner, 1990, 2006; Ali and Boylan, 2011, 2012).

When the corrected approach is used rather than the MSE approach, then the correction factor for the standard deviation of the lead time forecast error is

$$\begin{aligned}
 M_{\text{MSE}} &= \sqrt{\frac{L\sigma^2 + L^2\text{Var}(\hat{Y}_N)}{L\sigma^2 + L\text{Var}(\hat{Y}_N)}} \\
 &= \sqrt{\frac{1 + L\text{Var}(\hat{Y}_N)/\sigma^2}{1 + \text{Var}(\hat{Y}_N)/\sigma^2}} \\
 &= \sqrt{1 + \frac{(L-1)\text{Var}(\hat{Y}_N)/\sigma^2}{1 + \text{Var}(\hat{Y}_N)/\sigma^2}}.
 \end{aligned} \tag{2.3}$$

Similarly, compared to direct substitution of the demand variance  $\sigma^2$ , the correction factor for the standard deviation of the lead time forecast error is

$$M_\sigma = \sqrt{\frac{L\sigma^2 + L^2\text{Var}(\hat{Y}_N)}{L\sigma^2}} = \sqrt{1 + \frac{L\text{Var}(\hat{Y}_N)}{\sigma^2}}. \tag{2.4}$$

Observe that for  $L > 1$ , we have that  $M_\sigma > M_{MSE}$ . For normally distributed lead time demand, we have that the lead time forecast error is normally distributed with mean zero and variance according to (2.1). Furthermore, the safety stock for an  $(r, Q)$  policy is proportional to the standard deviation of the lead time forecast error for any cycle service level in the case of normally distributed demand (Zipkin, 2000; Axsäter, 2015). So, for normally distributed demand, (2.3) and (2.4) give the safety stock mark-ups of the correct approach relative to that resulting from the traditional approaches. Note from (2.3) and (2.4) that the correction factor increases with both the lead time and the relative variance of the forecast compared to the demand variance.

For SMA where the average is taken over  $M \leq N$  historical demands, we have

$$\frac{\text{Var}(\hat{Y}_N)}{\sigma^2} = \frac{\sigma^2/M}{\sigma^2} = \frac{1}{M}$$

and so the correction factor (2.3) becomes

$$M_{\text{MSE}, \text{SMA}} = \sqrt{1 + \frac{(L-1)/M}{1 + 1/M}} = \sqrt{1 + \frac{L-1}{M+1}}.$$

Similarly, mark-up (2.4) becomes

$$M_{\sigma, \text{SMA}} = \sqrt{1 + \frac{L}{M}}.$$

Both mark-up terms are increasing in  $L$  and decreasing in  $M$  as expected. For SES with smoothing constant  $\alpha$ , it is well-known (see e.g. Axsäter, 2006) that we have

$$\frac{\text{Var}(\hat{Y}_N)}{\sigma^2} = \frac{\alpha}{2 - \alpha} \tag{2.5}$$

and so the correction factor for the mark-up (2.3) becomes

$$M_{\text{MSE}, \text{SES}} = \sqrt{1 + \frac{(L-1)\alpha/(2-\alpha)}{1 + \alpha/(2-\alpha)}} = \sqrt{1 + \frac{(L-1)\alpha}{2}}.$$

Similarly, mark-up (2.4) becomes

$$M_{\sigma, \text{SES}} = \sqrt{1 + \frac{L\alpha}{2 - \alpha}}.$$

Both mark-up expressions are increasing in  $L$  and  $\alpha$ , as expected. Please note that the correction factors for SMA and SES are the same if  $M = \frac{2-\alpha}{\alpha}$ , which is also the condition for which the average age of the data in the forecasts is the same (Brown, 1963). For completeness, we remark that the finite-sample variance of the SES estimator is given by

$$\frac{\text{Var}(\hat{Y}_N)}{\sigma^2} = \frac{\alpha}{2 - \alpha} (1 - (1 - \alpha)^{2N}) + (1 - \alpha)^{2N}.$$

That expression should be used if the number of demand observations of the complete demand history,  $N$ , is relatively small. In the remainder of this paper we assume for SES that  $N$  is large enough so that (2.5) can be used.

In Tables 2.1 and 2.2, we show the percentage by which the traditionally calculated safety stocks need to increase for (normally distributed lead time demand and) SMA and SES, respectively, and for some typical control parameter values. Tables 2.1a and 2.2a show the mark-ups in case the one-period MSE is used to estimate the per-period demand variance, and Tables 2.1b and 2.2b depict the cases where the variance of the demand observations is directly substituted for  $\sigma^2$ .

It appears from Tables 2.1 and 2.2 that the increase in the lead time forecast error variance (and therefore in the safety stock that results from auto-correlation) is considerable for a wide range of control parameter values for both SMA and SES. For example, when forecasts are based on SMA over the last 12 periods, and the lead time is equal to 3 periods, then the traditional demand variance approach underestimates the safety stock level by 12%. When  $L = 1$ , there is obviously no auto-correlation of forecast errors, resulting in a 0% mark-up compared to the per-period MSE approach. The substantial mark-up percentages even for  $L = 1$  in Tables 2.1b and 2.2b show that directly estimating  $\sigma^2$ , and thereby ignoring the forecast error of mean demand, leads to safety stocks that are significantly too low even for short lead times. Naturally, the auto-correlation effect increases with the lead time. If  $M \rightarrow \infty$  or  $\alpha \rightarrow 0$  (which is what the traditional demand variance approach implicitly as-

sumes), the error of the forecast around the true mean of demand tends to 0 and the variance of the forecast error approaches the true demand variance, which leads to diminishing correction factors.

**Table 2.1:** Correction factors for safety stock calculation with an SMA procedure

**(a)  $L \times \text{MSE}$  approach**

	Increase in safety stock for			
$L$	$M = 1$	$M = 4$	$M = 12$	$M = 52$
1	0%	0%	0%	0%
2	22%	10%	4%	1%
3	41%	18%	7%	2%
4	58%	26%	11%	3%
5	73%	34%	14%	4%
6	87%	41%	18%	5%

**(b)  $L\sigma^2$  approach**

	Increase in safety stock for			
$L$	$M = 1$	$M = 4$	$M = 12$	$M = 52$
1	41%	12%	4%	1%
2	73%	22%	8%	2%
3	100%	32%	12%	3%
4	124%	41%	15%	4%
5	145%	50%	19%	5%
6	165%	58%	22%	6%

**Table 2.2:** Correction factors for safety stock calculation with an SES procedure

<b>(a) <math>L \times \text{MSE}</math> approach</b>				<b>(b) <math>L\sigma^2</math> approach</b>			
	Increase in safety stock for				Increase in safety stock for		
$L$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$L$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
1	0%	0%	0%	1	3%	5%	8%
2	2%	5%	7%	2	5%	11%	16%
3	5%	10%	14%	3	8%	15%	24%
4	7%	14%	20%	4	10%	20%	31%
5	10%	18%	26%	5	12%	25%	37%
6	12%	22%	32%	6	15%	29%	43%

## 2.5 Corrected demand distribution when both $\mu$ and $\sigma^2$ are unknown

The analysis up until now has been insightful in the sense that we could explicitly derive the lead time forecast error and indicate which parts are ignored by traditional approaches to calculate safety stocks. However, the results are still restrictive as the demand variance is typically unknown as well. In this section we focus on stationary, normally distributed demand and relax the assumption that the demand variance  $\sigma^2$  is known. That is,  $Y_t \stackrel{\text{i.i.d.}}{\sim} (\mu, \sigma^2)$  for  $t = 1, 2, \dots$  with both  $\mu$  and  $\sigma^2$  unknown. There is again an order lead time of length  $L > 0$ . As in Section 2.3, we derive the corrected lead time demand distribution in this setting. As mentioned in Section 2.2, Ritchken and Sankar (1984) treat inventory decision making under normally distributed demand with unknown parameters in a single-period model, implicitly using the average of all past demand observations as the estimator for the demand mean. We generalize their approach to the current setting with a positive lead time and to the SMA forecasting method.

Define the sample standard deviation

$$\hat{\sigma} = \sqrt{\frac{1}{M-1} \sum_{k=N-M+1}^N (Y_k - \hat{Y}_N)^2},$$

where  $M$  is the number of past observations based on which one would like to estimate the standard deviation of demand and  $\hat{Y}_N$  is the corresponding SMA estimator of the mean demand. Since  $\sum_{k=1}^L Y_{N+k} \sim N(L\mu, L\sigma^2)$  and, independently,  $L\hat{Y}_N \sim N(L\mu, L^2\text{Var}(\hat{Y}_N))$ , it follows that

$$\frac{\sum_{k=1}^L Y_{N+k} - L\hat{Y}_N}{\sqrt{L\sigma^2 + L^2\text{Var}(\hat{Y}_N)}} \sim N(0, 1).$$

Furthermore, by Cochran's theorem (Cochran, 1934) or Basu's theorem (Basu, 1955)

it follows that, independently,

$$\frac{(M-1)\hat{\sigma}^2}{\sigma^2} \sim \chi_{M-1}^2,$$

where  $\chi_{M-1}^2$  denotes the chi-square distribution with  $M-1$  degrees of freedom. For  $A \sim N(0, 1)$  and  $B \sim \chi_{\beta}^2$ , with  $A$  and  $B$  being independent, it holds that

$$\frac{A}{\sqrt{B/\beta}} \sim t_{\beta},$$

where  $t_{\beta}$  denotes the Student's  $t$ -distribution with  $\beta$  degrees of freedom. When we substitute

$$\begin{aligned} A &= \frac{\sum_{k=1}^L Y_{N+k} - L\hat{Y}_N}{\sqrt{L\sigma^2 + L^2\text{Var}(\hat{Y}_N)}}, \\ B &= \frac{(M-1)\hat{\sigma}^2}{\sigma^2}, \end{aligned}$$

and  $\beta = M-1$ , we find that

$$\frac{\sum_{k=1}^L Y_{N+k} - L\hat{Y}_N}{\sqrt{L\hat{\sigma}^2 + L^2\widehat{\text{Var}}(\hat{Y}_N)}} \sim t_{M-1}, \quad (2.6)$$

where  $\widehat{\text{Var}}(\hat{Y}_N) = \hat{\sigma}^2/M$  for SMA. This specifies the corrected lead time demand distribution.

Observe that this closed-form distribution can only be obtained for an SMA estimator (or its special case, the average over all demand observations). Basu (1955) discusses that two statistics are independently distributed if the first statistic is a sufficient statistic for a certain parameter and the distribution of the other statistic does not depend on that parameter. In this case the SMA estimator is a sufficient statistic for  $\mu$  in the restricted sample consisting of the last  $M$  demand observations, and the distribution of the sample variance does not depend on  $\mu$ . However, the SES estimator is a special case of a weighted average estimator and is therefore not a sufficient

statistic for  $\mu$  (as it is not a direct transformation of the sum of all demand observations). Therefore, the closed-form results based on the Student's t-distribution cannot be extended to the SES forecasting technique.

In case an SES forecasting technique is used, one can simulate the joint distribution of the SES estimator for the mean and the sample variance and search for the correct reorder level. However, given that the stationary demand model is correctly specified, the same reorder level will be found as when SMA is used on the same demand data set, as this is the single reorder level that achieves the desired service level. This indicates that the appropriate forecasting method is implied by the demand model, and that forecasting and decision making should be integrated.

## 2.6 Applications to inventory systems

In this section, we will use the different approaches to express the lead time forecast error variance to find the reorder levels for inventory systems with an  $(r, Q)$  replenishment policy. A target cycle service level  $\gamma^*$  is used to determine the reorder level  $r$ . This service level is defined as the probability that an order arrives on time, i.e. that no stock-out occurs in a replenishment cycle. The advantage of this service level measure is that it does not depend on the batch size  $Q$ .

Since demand is normally distributed, the reorder levels for an  $(r, Q)$  policy according to the traditional methods (direct substitution and the MSE approach), and the corrected method presented in this paper, are easily characterized by (2.7), (2.8), and (2.9), respectively:

$$r_\sigma = L\hat{Y}_N + \Phi^{-1}(\gamma^*)\sqrt{L}\sigma, \quad (2.7)$$

$$r_{\text{MSE}} = L\hat{Y}_N + \Phi^{-1}(\gamma^*)\sqrt{L\sigma^2 + L\text{Var}(\hat{Y}_N)}, \quad (2.8)$$

$$r^* = L\hat{Y}_N + \Phi^{-1}(\gamma^*)\sqrt{L\sigma^2 + L^2\text{Var}(\hat{Y}_N)}, \quad (2.9)$$

where  $\Phi$  denotes the standard normal distribution function,  $\text{Var}(\hat{Y}_N) = \sigma^2/M$  if  $\hat{Y}_N$  is based on SMA over  $M$  observations, and  $\text{Var}(\hat{Y}_N) = \sigma^2\alpha/(2 - \alpha)$  if  $\hat{Y}_N$  is based on SES with smoothing constant  $\alpha$ . Applying reorder level  $r$ , the actually achieved cycle service level  $\gamma'$  based on the true demand parameterization is easily computed



by

$$\gamma' = \Phi\left(\frac{r - L\mu}{\sqrt{L\sigma^2}}\right).$$

If we furthermore relax the assumption that  $\sigma$  is known, then we can proceed analogously for the SMA forecasting technique. Based on the (observed) estimates  $\hat{Y}_N$  and  $\hat{\sigma}$ , we can select the reorder level  $r$  so that

$$\mathbb{E}\left\{\mathbb{P}\left(r - \sum_{k=1}^L Y_{N+k} \geq 0\right)\right\} = \gamma^*.$$

Let  $X$  be a random variable following the Student's t-distribution with  $M - 1$  degrees of freedom. We have

$$\begin{aligned} \mathbb{P}\left(r - \sum_{k=1}^L Y_{N+k} \geq 0\right) &= \mathbb{P}\left(\sum_{k=1}^L Y_{N+k} \leq r\right) \\ &= \mathbb{P}\left(X \leq \frac{r - L\hat{Y}_N}{\sqrt{L\hat{\sigma}^2 + L^2\widehat{\text{Var}}(\hat{Y}_N)}}\right). \end{aligned}$$

This implies that the optimal reorder level is given by

$$r^* = L\hat{Y}_N + t_{M-1:\gamma^*}\sqrt{L\hat{\sigma}^2 + L^2\widehat{\text{Var}}(\hat{Y}_N)},$$

where  $t_{M-1:\gamma^*}$  is the  $\gamma^*$ th percentile of the Student's t-distribution with  $M - 1$  degrees of freedom and  $\widehat{\text{Var}}(\hat{Y}_N) = \hat{\sigma}^2/M$  for SMA over  $M$  observations. Observe the analogy with the reorder level (2.9) which assumes that  $\sigma$  is known, and therefore uses a quantile from the standard normal distribution rather than the Student's t-distribution.

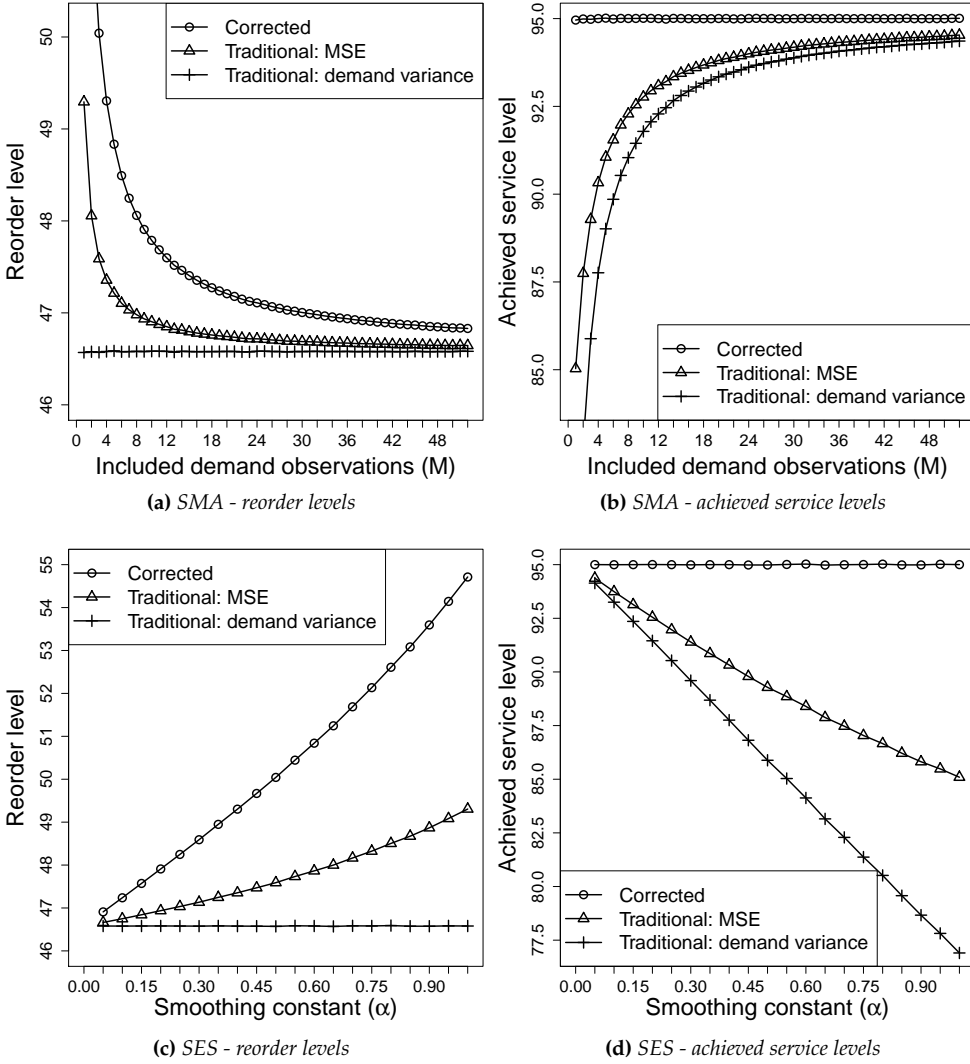
## 2.7 Numerical results

In this section, we evaluate the significance of the error that results from traditional approaches to set the reorder levels by means of several numerical simulation examples. We first discuss the setting where the mean of the normally distributed demand

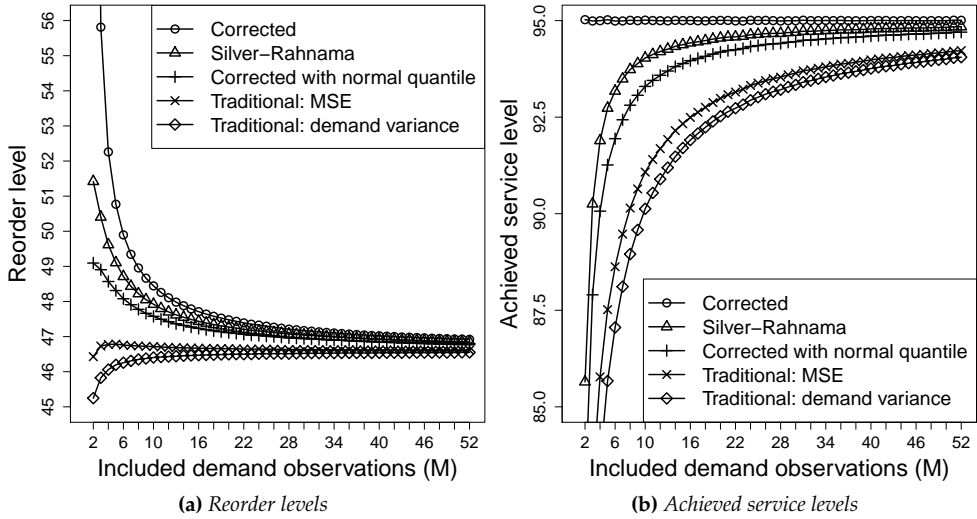
is assumed unknown, whereas the variance is known. For both the SMA and SES forecasting techniques we calculate the reorder levels for the  $(r, Q)$  policy according to the demand variance approach, MSE approach, and the corrected approach for different choices of the SMA number of observations  $M$  and the SES smoothing constant  $\alpha$ . In all settings we select the mean of per-period demand  $\mu = 10$ , the standard deviation  $\sigma = 2$ , the fixed lead time  $L = 4$ , and the target service level  $\gamma^* = 95\%$ . For SMA we let the number of included demand observations  $M$  range from 1 to 52 and for SES we let the smoothing constant  $\alpha$  range from 0.05 to 1. Figure 2.1 presents the reorder levels  $r$  and the corresponding service levels for the traditional and corrected approaches.

First consider SMA. The reorder level correction that is needed compared to the demand variance approach increases up to almost 20%. Especially when  $M$  increases, the reorder levels corresponding to the MSE and demand variance approach quickly converge towards each other, whereas a gap remains with the corrected reorder level. This is also reflected in the achieved service levels. The corrected approach achieves its target for all scenarios, but the demand variance approach undershoots this target by up to 17.5%. This implies more than three times as many stock-outs as were intended. The MSE approach is off by up to 10%. For  $M = 30$  both traditional approaches still show a gap with the corrected approach of approximately 2%, and for  $M = 52$  this gap is approximately 1%. If the moving average is taken over, for example, 8 past demands, then the MSE approach leads to a service level of 92% instead of 95%, implying 60% more stock-outs. For the demand variance approach the service level drops to 89% for  $M = 8$ . We conclude that the results achieved by the traditional approaches are not satisfactory even if relatively many data points are available.

For SES the results are similar. Especially for large values of the smoothing constant  $\alpha$  the traditional approaches lead to too low reorder levels and service levels. For these large values, the average age of the data is small, which means that the estimation error is large. Furthermore, observe that the service level difference between the corrected approach and the demand variance approach increases exponentially as  $M$  decreases for SMA, and linearly as  $\alpha$  increases for SES. The reorder levels and service levels are the same for SMA and SES if  $M = \frac{2-\alpha}{\alpha}$ .



**Figure 2.1:** Reorder levels and achieved service levels for the corrected approach and the traditional MSE and demand variance approaches, for SMA and SES. Normally distributed demand with  $\mu = 10$  (unknown),  $\sigma = 2$  (known),  $L = 4$  (known), and target service level  $\gamma^* = 95\%$ . Results based on 1,000,000 repetitions.



**Figure 2.2:** Reorder levels and achieved service levels for the corrected approach based on the Student's  $t$ -quantile, the approach based on the normal quantile, and the traditional MSE and demand variance approaches. Normally distributed demand with  $\mu = 10$  (unknown),  $\sigma = 2$  (unknown),  $L = 4$  (known), target service level  $\gamma^* = 95\%$ , and an SMA forecasting procedure. Results based on 1,000,000 repetitions.

Figure 2.2 shows the results for the SMA forecasting technique under the assumption that the variance of the demand is unknown. Besides the comparison between the corrected approach and the traditional demand variance and MSE approaches, we also included the results from the corrected approach when the demand variance is known. We furthermore show the reorder levels that result from the procedure of Silver and Rahnama (1987). The number of included demand observations range from 2 to 52. At least 2 observations are needed to estimate both the mean and standard deviation of the demand.

The discrepancy between the traditional methods and the corrected method is larger in this scenario, as the estimation uncertainty of  $\sigma$  is ignored by the traditional approaches. For small values of  $M$ , the reorder level difference between the corrected approach and the traditional demand variance approach increases up to 67%, implying that the safety stock difference is 600%, for the extreme case that  $M = 2$ . As  $M$  increases, the difference gradually decreases. The method proposed by Silver and Rahnama (1987) performs significantly better than the traditional approaches,

but still sets the reorder level too low. The corrected expression that only ignores uncertainty of the demand variance also performs better than the traditional approaches. Furthermore, the service level differences between the different methods are larger than in the scenario with  $\sigma$  known. Observe that, counter-intuitively, the reorder levels for the traditional approaches slightly increase as  $M$  increases. That is, basing the forecast on more observations leads to a (somewhat) higher reorder level. This is a result of the sample standard deviation being a biased estimator of  $\sigma$ .

## 2.8 Conclusion

We argued that the issue of not knowing the mean and variance of the demand distribution has largely been ignored in both inventory theory and practice, and that using either the one period ahead forecast error or the variance of per-period demand provides results that are far from optimal when a positive lead time is present. This will always lead to a safety stock that is smaller than what is needed to meet a prescribed target service level, and in many cases it will be far too small. Also the method proposed by Silver and Rahnema (1987), though computationally heavy, does not achieve the target service level.

Simple, closed-form expressions were derived for correcting the lead time forecast error variance and (for normally distributed demand) the corresponding safety stocks. It transpired that even when the demand process is not auto-correlated, the future forecast errors are correlated, and the traditional approaches, even when based on the minimum variance unbiased estimator, lead to considerable errors for a wide range of forecasting control parameter settings.

The approach taken in this paper can be extended to other types of inventory models, also if multiple parameters are to be set. One should determine at which points the decision depends on the demand distribution and substitute the corrected distribution for the unknown true demand distribution. This holds in simple models that can be solved exactly, such as the  $(r, Q)$  model with a fixed order quantity  $Q$ , but also in more complex models. Furthermore, the analysis can be extended to other demand processes and forecasting procedures, although analyzing non-stationary demand processes is considerably more complex. Another direction for future research is to consider systems with stochastic lead times. As such systems are notoriously

difficult to analyze even without the consideration of forecast errors, assuming the lead time distribution to be given may be appropriate to obtain initial insights. In practice, the lead time distribution will have to be estimated (as well), and inclusion of these forecast errors should also be considered.

A final interesting route to explore is taking the lead time as the forecasting period. By doing so, the lead time reduces to a single period and therefore the issue of auto-correlated forecast errors over multiple periods is avoided. However, this approach has a number of disadvantages. First, different forecasting periods are needed for different stock keeping units. Second, situations with stochastic lead times cannot be dealt with in this way. Third, and more technical, calculating the so-called undershoot in inventory control systems (i.e. the amount by which the stock position drops below the reorder level) is impossible using aggregated demand information.

Given the considerable discrepancies that were observed under stationary demand, and the fact that the underlying issue of correlated forecast errors is present for any demand process and any forecasting procedure, wider recognition and further exploration are certainly needed. Moreover, developers and users of forecasting and inventory control software should be made aware of the pitfalls of traditional safety stock rules.

